DOCUMENT RESUME

ED 065 532 TM 001 453

AUTHOR	Evans, Franklin R.; Reilly, Richard R.
TITLE	A Study of Test Speededness as a Potential Source of Bias in the Admission Test for Graduate Study in Business Quantitative Score.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-72-9
PUB DATE	Feb 72
NOTE	24p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Performance; *Admission Criteria; *Aptitude Tests; Business Education; Business Skills; Caucasians; Ethnic Groups; Evaluation Techniques; *Graduate Study; *Minority Groups; National Competency Tests; Negroes; *Test Bias; Test Reliability; *Timed Tests
IDENTIFIERS	Admission Test for Graduate Study in Business; ATGSB

ABSTRACT

Minority and majority groups were administered a special quantitative section of the Admission Test for Graduate Study in Business (ATGSB) under varying time conditions to determine if increasing the time allotted for the test would eliminate any bias which may exist due to an irrelevant speed factor. By a commonly employed definition the special section was found to be moderately speeded for all candidates under normal conditions. Neither the main effects due to time condition nor the interaction between the ethnic and time factor reached significant levels suggesting that increasing the time per item does not reduce any bias which may exist in the test. Although a substantial proportion of minority group scores fell at or below the chance level, these scores appeared to retain fairly high levels of reliability. (Author)

ED 065532

RESEARCH BULLETIN

TM 001 453

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RB-72-9

A STUDY OF TEST SPEEDEDNESS AS A POTENTIAL SOURCE OF BIAS IN THE ADMISSION TEST FOR GRADUATE STUDY IN BUSINESS QUANTITATIVE SCORE

Frankl.n R. Evans

and

Richard R. Reilly

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
February 1972

A STUDY OF TEST SPEEDEDNESS AS A POTENTIAL SOURCE OF BIAS IN THE ADMISSION TEST FOR GRADUATE STUDY IN BUSINESS QUANTITATIVE SCORE

Franklin R. Evans and Richard R. Reilly

Abstract

Minority and majority groups were administered a special quantitative section of the Admission Test for Graduate Study in Business (ATGSB) under varying time conditions to determine if increasing the time allotted for the test would eliminate any bias which may exist due to an irrelevant speed factor. By a commonly employed definition the special section was found to be moderately speeded for all candidates under normal time conditions. Neither the main effects due to time condition nor the interaction between the ethnic and time factor reached significant levels suggesting that increasing the time per item does not reduce any bias which may exist in the test. Although a substantial proportion of minority group scores fell at or below the chance level, these scores appeared to retain fairly high levels of reliability.

A STUDY OF TEST SPEEDEDNESS AS A POTENTIAL SOURCE OF BIAS IN THE ADMISSION TEST FOR GRADUATE STUDY IN BUSINESS QUANTITATIVE SCORE[1]

Franklin R. Evans and Richard R. Reilly
Educational Testing Service

Standardized academic aptitude tests have been the subject of persistent criticism from members of certain minority groups who charge that such tests are unfair to members of their groups. Flaugher (1970) in a recent review of testing practices with respect to minority groups discussed three potential sources of unfairness which may be summarized as: (a) those having to do with test content, (b) the conditions or circumstances under which standardized tests are administered, and (c) the way in which test scores are actually used. Much recent research has centered on the third possible source of bias, with most researchers considering a test unbiased if the regression of the criterion scores on the test is the same for both groups. Thorndike (1971) has demonstrated that the use of a test that is unbiased by this definition will result in the screening out of a larger proportion of the minority group candidates than would be the case if the test were perfectly valid. Thorndike also pointed out that, "...one cannot appraise the 'fairness' of a test through its correlation with an 'unfair' criterion."

In the absence of specific knowledge about how much bias, if any, exists in the criterion, the second possible source of bias mentioned above, the conditions under which tests are administered might be profitably investigated as possible biasing factors. One common complaint about test administration conditions made by spokesmen from culturally disadvantaged groups is that standardized academic aptitude tests are too highly speeded. Almost all academic aptitude and achievement tests purport to be primarily measures of "power" and not speed: i.e.,

most candidates are expected to have the opportunity to attempt all or almost all of the items in the allotted time. A recent study by Evans and Reilly (1972) investigated the effects of varying speededness of a reading comprehension section of the Law School Admission Test (LSAT) on minority versus majority group performance. Although they report that lowering the degree of speededness in a reading comprehension section of the LSAT did not benefit one group more than the other (in terms of gain in mean score level), it was concluded that this section was more speeded for minority than for majority group candidates. The purpose of the present investigation was to extend these results to a test with a lower verbal "load."

The Admission Test for Graduate Study in Business (ATGSB) is designed to yield a quantitative as well as a verbal score. The present study was conducted with the intent of determining: (1) if the quantitative section of the ATGSB is more speeded for Black examinees than for White, and (2) if reducing or increasing the degree of speededness has a differential effect on the scores of minority versus majority group members.

Procedure

Subjects

The ATGSB is a nationally administered test, and data for the present study were collected from a regularly scheduled ATGSB administration in 1971. In addition, data were collected from candidates taking the ATGSB at 26 centers in predominantly Black colleges in the Southeastern United States where no test fee was charged. A special research section was included with the five operational sections of the ATGSB, and three different versions of this special section were administered by spiralling forms so that roughly one-third of the candidates in the study took each of the special forms.

Regular center Black candidates (RCB) and regular center White candidates (RCW) were identified within the national sample by means of a set of background questions. This same set of questions was administered at the special centers, but since the overwhelming majority of candidates at those centers was Black, only one special group, special center Black (SCB), was identified for purposes of the study.

Experimental Subtests

In order to investigate the research questions posed above it was necessary to create experimental forms of a test which differed only in the degree to which they were speeded. The most obvious way to accomplish this would be to administer identical tests under different time conditions. Because of the constraints imposed by the standardization necessary for a national testing program, however, it was necessary to find an alternative method of varying speed. In this study, three different speed conditions were created by varying the number of five-choice mathematics items within the special section while holding the time limit constant. Although all candidates took the special section under the 40-minute time limit, Form A of the special section had 25 items; Form B, 30; and Form C, 35. Forms B and C were among a set of 25 items common to all forms. Scores on these 25 items served as the dependent variable for most of the analyses in the present study. Table 1 shows the order of the items in the three special sections. The net effect of this was that rather than the absolute time allowed

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

being varied, the amount of time per item was changed. Under normal circumstances the average time per item on the ATGSB quantitative section is about 82 seconds. This may be compared with the averages of 96 seconds for Form A, 80 seconds for Form B, and 69 seconds for Form C. Thus, the time conditions under which the

special section was administered included one "normal" condition (Form B), one "speeded" condition (Form C), and one "unspeeded" condition (Form A).

Study Design

Because the candidates taking the ATGSB at special centers were not typical of candidates in general, it was decided to keep the RCB and SCB groups separate for purposes of analysis. Thus, the study design was a 3 x 3 in which specific attention was focused on the interaction between time condition and ethnic/center group. That is, there was less interest in the main effects due to test forms and less still in the main effects due to the group factor than there was in the possible differential effects due to speededness among the three groups.

Results and Discussion

- -

Insert Figures 1a, 1b, and 1c about here

- -

The first question this study attempted to answer was whether the ATGSB quantitative test is more speeded for Blacks than for Whites. According to criteria used by Swineford (1956) a test may be considered unspeeded if (1) virtually all candidates reach 75 per cent of the items and (2) at least 80 per cent of the candidates respond to the last item. As Figures 1a,b, and c show the special section appeared to be a speeded test under all three time conditions for all three groups.[2] The items in the special section were both ordered in difficulty (from easy to difficult) and corrected for guessing, which means that many individuals may not have attempted the last few items, not because they ran out of time, but because the items were simply too

difficult and they chose not to guess. It is clear from examining the figur that a substantial proportion of dropouts (20 per cent) occurred, in general somewhat earlier for the Black group than for the White group. The differences observed in dropout rate in this study were not nearly as striking as those reported in an earlier study of a test in which the items were neither ordered in difficulty nor corrected for guessing (Evans & Reilly, 1972). It is also worth noting that unlike that study, no clear relationship among the mean scores of the various groups and the rate at which members of these groups did not complete the test was found. (Table 2 shows mean score levels[3] for the three different groups taking each test form.) In fact, under the

- - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - -

more speeded conditions (Forms B and C) the group with the lowest mean scores (SCB) actually exhibited the lowest dropout rate.

The next question the study attempted to answer was whether raising or lowering the degree of speededness of the quantitative section of the ATGSB would differentially affect scores among the three ethnic/center groups identified, and, in particular, whether increasing the amount of time per item would more greatly benefit the minority groups. One possible approach would have been to examine the interaction effect between the time condition and group factors in a 3^2 factorial analysis of variance, but the authors chose to analyze the data in a slightly different way in an attempt to reduce some of the error due to a less than optimum procedure for assigning individuals to treatment (time condition). It was clear that scores on a regular quantitative section of the ATGSB would be highly correlated with scores on the special section, but it was equally clear that some of the necessary assumptions for using that score as a covariate in an analysis of covariance could not be met. Although individuals were assigned

approximately randomly within ethnic/center group, there were rather large differences (as can be seen in Table 3) in mean scores on the covariate across the three ethnic/center groups and there was no guarantee that the same regression line could be used to "adjust" scores for all three groups. Cautions against

- - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - -

using analysis of covariance in this type of situation have become almost cliché (e.g., Evans & Anastasio, 1968; Lord, 1967). In this study primary interest was on the interaction effect, and since assignment within group was approximately random, a slightly different linear model was used in an attempt to overcome the aforementioned shortcomings of traditional analysis of covariance. The following model was used:

$$y_{ijk} = \alpha_i + \rho_i x_{ijk} + \beta_j + \gamma_{ij} + e_{ijk}$$

where x and y scores are expressed as deviations from their respective grand means; ρ_i and α_i represent the within-group slopes and intercepts respectively; x_{ijk} is a covariate; β_j is the main effect due to time condition; γ_{ij} is the interaction between group and time condition; and e_{ijk} is the error term.

This model actually uses separate regression lines derived from data across treatments within a given ethnic/center group, and in effect, the dependent variable becomes the deviation of the special section score about the within-group regression line. Obviously, any main effects due to ethnic/center group differences could not be tested with such a model. In the present study, however, there was no interest in testing group differences, and the use of separate within-group regression lines should have had the effect of merely reducing the error term to be used in testing the hypotheses of interest, that is, the main effects due to time condition, and especially the interaction effect. It may be

of interest to some readers to note that this analysis was performed using a standard multiple regression computer program. The results of the analysis, for comparative purposes, are presented in Table 4 along with the results of an analysis of variance performed on the same data.

- - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - -

Neither of the tests of interest reached significance levels, and the proportions of variance accounted for by the time condition and the interaction effects, respectively, were both near zero in the analysis of variance. Thus, raising or lowering the time per item appears to have had almost no effect in changing score levels and no differential effect among ethnic/center groups. The results of this study appear even more conclusive than the results of an earlier study (Evans & Reilly, 1972) in the verbal domain, where a significant main effect due to time per item was observed and a slight but not significantly beneficial effect for Black candidates resulted when the time per item was increased. Aside from the possibility that quantitative abilities are more resistant to speed effects than verbal abilities, the major reason for this discrepancy in results may be the way in which the experimental sections in each of the studies were constructed. In the previous study, the items were not clearly ordered for difficulty and there was no correction for guessing, while the reverse was true for the experimental section in the present investigation. Both ordering items in terms of their difficulties and correcting for guessing would tend to weaken any effects due to speed.

- - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - -

Kuder-Richardson (KR-20) reliabilities were computed for each group under each time condition and are presented in Table 5 along with the correlations between the special section and another quantitative section (Q) which may be regarded as approximating parallel forms reliabilities.

It should be noted that section Q could not be considered truly parallel since Q differs in number of items (55 for Q vs. 25 for the special section) and was administered under invariant time conditions. However, the special section was constructed in such a way as to approximate the item characteristics of the regular quantitative section of the ATGSB. The distributions of item difficulties and item total correlations were approximately the same. The "parallel-form" reliabilities reported in the lower half of Table 5 were corrected to account for the differing number of items by use of the formula

$$r_{xx'} = r_{xy} \frac{N_x}{N_y} \cdot \frac{\sigma_y}{\sigma_x},$$

where r_{xy} is the correlation between the two forms

N_x is the number of items in the shorter of the forms

N_y is the number of items in the longer of the forms

σ_x and σ_y the standard deviations of the forms.

As can be seen in Table 5 the K-R coefficient tends to underestimate the parallel forms reliability under less speeded conditions (Form A) and to overestimate this coefficient under more speeded conditions (Form C). This pattern is rather interesting but the differences are not in general very large, especially in the two regular center groups. However, in the SCB group it may be noted that the single form reliability (K-R) overestimates the "parallel forms" reliability by a large amount for all three forms possibly reflecting the presence of a large speed component for all three SCB groups.

Speed may have had the effects of increasing the proportion of error variance for the SCB group, thereby spuriously increasing the estimate of reliability. As mentioned earlier the dropout rates shown in Figures 1a,b, and c probably do not reflect accurately the effects of speed in a test which is both ordered for difficulty and corrected for chance success.

It is also worth noting that even though in two of the groups (SCB and RCB) a substantial portion of the special sections score range was near or below the chance level, the scores remained fairly reliable. The reliability and predictive value of scores in the chance range has been reported in the literature (Boldt, 1968) and is of interest, since in order to obtain reliable less-than-chance scores individuals have to be operating under a decidedly less than optimal guessing strategy.

One of the criticisms leveled at standardized tests by minority group spokesmen has been that performance on such tests is to some degree dependent on a "test-wiseness" not possessed by minority group members. The ability to handle guessing instructions appropriately and adopt optimum or near optimum guessing strategies is certainly related to test-wiseness, and it may be that the relative unfamiliarity of some minority group members with guessing instruction tends to put them at a slight disadvantage. It should be mentioned, however, that most national academic testing programs do not, in fact, report negative scores as such but rather scale below-chance scores upward to the lowest positive standardized score. Individuals receiving negative scores on the ATGSB, for example, would in practice have a reported score of 200. Research should probably be done,

nevertheless, to answer some of the questions surrounding test-talking strategies as they relate to minority groups.

Conclusions

The ATGSB quantitative section appeared to be a moderately speeded measure for both majority and minority group candidates, but neither increasing nor decreasing the time per item appeared to result in any differential effect among the three ethnic/center groups included in the study, suggesting that such a procedure would not eliminate any bias possibly due to speededness.

References

Boldt, R. F. Study of linearity and homoscedasticity of test scores in the chance range. Educational and Psychological Measurement, 1968, 28, 47-60.

Evans, F. R., & Reilly, R. R. A study of speededness as a source of test bias. Journal of Educational Measurement, 1972 (in press).

Evans, S. H., & Anastasio, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 1968, 69, 225-234.

Flaugher, R. L. Testing practices, minority groups, and higher education: A review and discussion of the research. Research Bulletin 70-41. Princeton, N. J.: Educational Testing Service, 1970.

Lord, F. M. A paradox in the interpretation of group comparisons. Psychological Bulletin, 1967, 68, 304-305.

Swineford, F. Technical manual for users of test analyses. Statistical Report 56-42. Princeton, N. J.: Educational Testing Service, 1956.

Thorndike, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8 (2), 63-70.

Footnotes

[1]The authors wish to thank The Graduate Business Admissions Council which supported this research.

[2]More detailed item data are presented in the Appendix.

[3]Unless otherwise noted the dependent variable or "score" referred to will be the corrected score on the 25 items common to Forms A, B, and C; i.e., the sum of the rights minus 1/4 the sum of the wrongs. All scores discussed are corrected scores.

[4]Random sampling procedures were used to create an orthogonal design by sampling down to the smallest cell size.

[5]The R^2 values are the squared multiple correlations with the dummy variables representing each factor indicated removed.

Table 1

Order of Items for Table 1, Forms A, B, C[a]

	Order of Items
Form A (25 Items)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Form B (30 Items)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Form C (35 Items)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

[a]The items are numbered separately for each form of the test. Items in the same column are identical. The blank spaces in Forms A and B indicate items which do not appear on these forms. For example, the first two items are the same on all three forms, the third item on Form C does not appear on Forms A and B, the sixth item on Form C is the same as the fifth item on Form B and does not appear on Form A.

Table 2

Descriptive Statistics for Groups in the Study Sample

(25 Item Corrected Scores)

		Form A	Form B	Form C
RCB	N	89	96	70
	Mean	6.59	7.42	6.16
	S.D.	4.90	4.97	4.79
	Range	-3.75 to 18.75	-5.0 to 22.5	-5.0 to 18.75
	% ≦ 0[a]	21.2	15.6	35.7
RCW	N	432	435	464
	Mean	13.79	13.08	12.30
	S.D.	5.26	5.09	7.70
	Range	-2.5 to 25	-6.25 to 25	-6.25 to 25
	% ≦ 0	1.8	1.6	2.5
SCB	N	245	249	239
	Mean	4.21	4.04	3.46
	S.D.	3.79	3.47	3.45
	Range	-3.25 to 17.5	-1.75 to 16.75	-4.75 to 13.75
	% ≦ 0	13.5	14.1	19.7

[a] % ≦ 0 indicates the percentage of candidates whose corrected scores were zero or less.

Table 3

Mean Scores for Section Q and Correlations

between Section Q and Special Section (S)

Group	Time Condition								
	Unspeeded			Normal			Speeded		
	N	$\bar{X}_q$	r_{qs}	N	$\bar{X}_q$	r_{qs}	N	$\bar{X}_q$	r_{qs}
RCB	89	12.76	.82	96	14.01	.79	70	13.48	.84
RCW	432	26.85	.81	435	27.09	.79	464	26.66	.79
SCB	245	6.71	.65	249	6.49	.70	239	6.01	.63

Table 4

Summary of Analysis of Covariance and Analysis of Variance

Factor	Covariance Analysis			
	R^2	d.f.[4]	F	
Ethnic/Center Group	.7758	2/627	3.51*	
Test Form	.7771	2/627	1.69	
Group X Form	.7769	4/627	.64	
All Variables	.7783			
	Analysis of Variance			Proportion of Variance Accounted for (R^2)[5]
Ethnic/Center Group	.2670	2/630	66.65*	.39
Test Form	.3921	2/630	1.61	.01
Group X Form	.3948	4/630	.10	.00
All Variables	.3952			

*$p < .05$

Table 5

Kuder-Richardson and "Parallel Forms" Reliabilities

for Ethnic/Center Group and Test Form

Group	Kuder-Richardson Coefficients		
	Form A	Form B	Form C
RCB	.79	.81	.83
RCW	.79	.80	.81
SCB	.82	.82	.84
	"Parallel Forms" Reliabilities		
RCB	.70	.74	.80
RCW	.71	.69	.76
SCB	.55	.67	.51

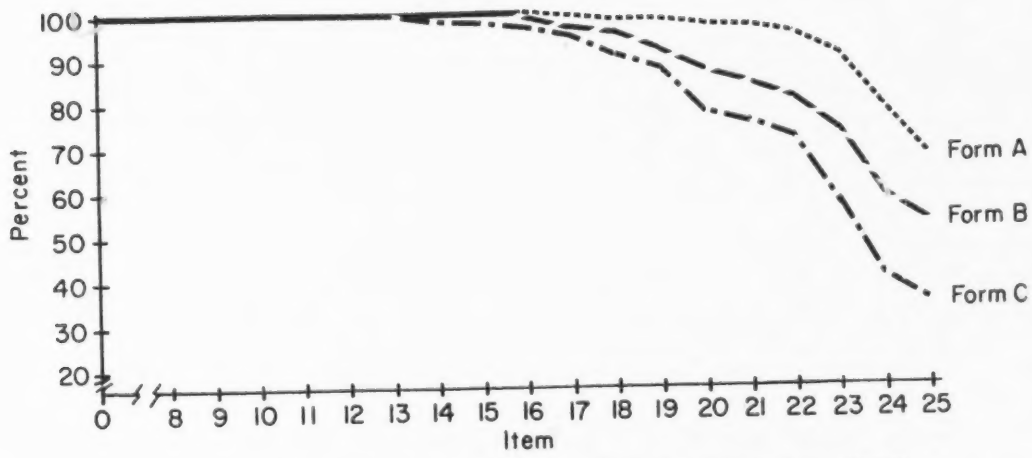

Figure Ia. Percent Items Reached for White Regular Center (WRC)

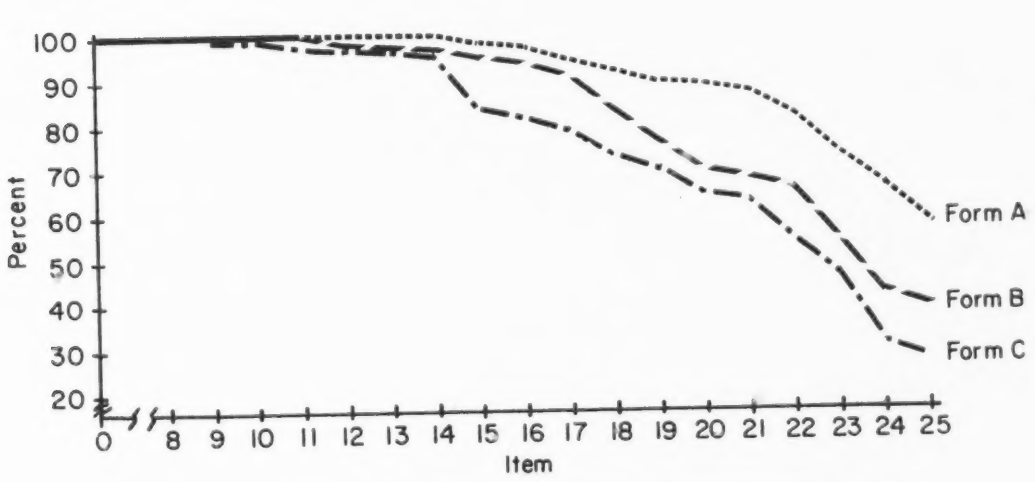

Figure Ib. Percent Items Reached for Black Regular Center (BRC)

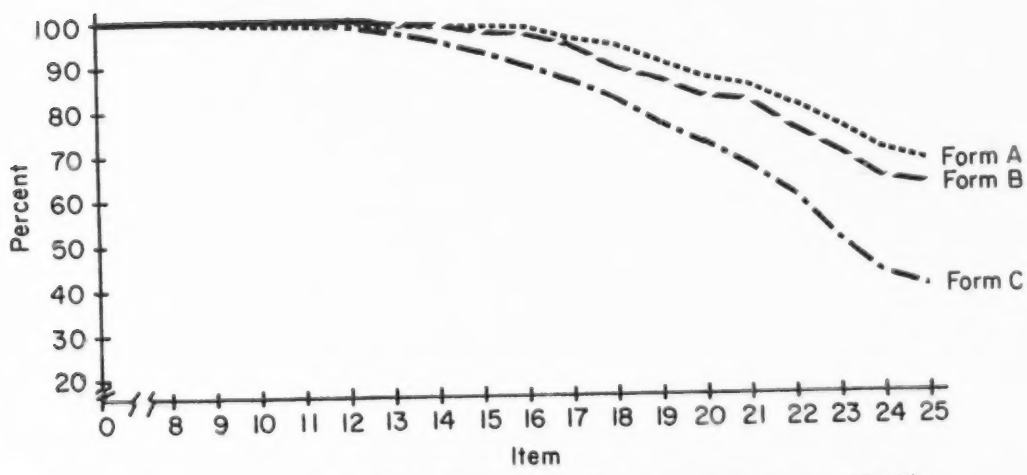

Figure Ic. Percent Items Reached for Black Special Center (BSC)

Appendix A

Items Deltas and Percentages of Attempts Not Reached and Correct Responses for Items Common to Forms A, B, and C

Form A

Item	RCB (N = 89) % Correct*	Delta**	% Attempt.	% Not*** Reached	RCW (N = 432) % Correct	Delta	% Attempt.	% Not Reached	SCB (N = 245) % Correct	Delta	% Attempt.	% Not Reached
1	72%	10.68	98%	0	86%	8.68	99%	0	63%	11.68	96%	0
2	77%	10.04	92%	0	87%	8.48	98%	0	72%	10.68	83%	0
3	53%	12.68	81%	0	78%	9.92	94%	0	37%	14.32	76%	0
4	60%	12.00	91%	0	86%	8.68	97%	0	59%	12.08	82%	0
5	69%	11.00	83%	0	87%	8.48	96%	0	67%	11.24	87%	0
6	71%	10.80	79%	0	91%	7.64	88%	0	67%	11.24	80%	0
7	49%	13.12	94%	0	80%	9.64	97%	0	25%	15.68	93%	0
8	71%	10.80	65%	0	92%	7.36	82%	0	61%	11.88	69%	0
9	25%	15.68	75%	0	64%	11.56	83%	0	13%	17.52	82%	1%
10	61%	11.88	79%	0	88%	8.32	92%	0	54%	12.60	73%	1%
11	49%	13.12	76%	0	73%	10.56	90%	0	46%	13.40	78%	1%
12	49%	13.12	76%	0	74%	10.44	92%	0	32%	14.88	81%	1%
13	28%	15.32	65%	0	68%	11.12	70%	0	22%	16.08	64%	1%
14	43%	13.72	91%	0	65%	11.44	93%	0	30%	15.08	84%	2%
15	50%	13.00	63%	2%	78%	9.92	72%	0	35%	14.56	59%	2%
16	69%	11.00	79%	3%	88%	8.32	90%	0	48%	13.20	77%	2%
17	37%	14.32	61%	6%	73%	10.56	85%	0	30%	15.08	69%	5%
18	44%	13.60	70%	8%	76%	10.16	78%	1%	29%	15.20	75%	6%
19	53%	12.68	76%	11%	73%	10.56	90%	1%	41%	13.92	75%	10%
20	40%	14.00	47%	12%	62%	11.80	66%	3%	25%	15.68	51%	14%
21	19%	16.52	66%	13%	46%	13.40	75%	3%	9%	18.36	71%	16%
22	22%	16.08	71%	18%	37%	14.32	90%	5%	13%	17.52	64%	20%
23	12%	17.68	46%	27%	38%	14.20	84%	9%	10%	18.12	44%	25%
24	45%	13.52	35%	34%	60%	12.00	61%	21%	20%	16.36	33%	30%
25	0	---	58%	42%	18%	16.68	68%	32%	5%	19.58	67%	33%

* % Correct is based on number attempting each item.

**Delta is the normal deviate, expressed in terms of a scale with mean of 13 and standard deviation of 4, which corresponds to the proportion of candidates attempting the item who answer it correctly.

***Not reached is the percent of people who did not attempt each item and did not attempt any subsequent items. For the last item it is simply the percent of people who did not respond to the item.

Form B

Item	RCB (N = 96) % Correct*	Delta**	% Attempt.	% Not*** Reached	RCW (N = 435) % Correct	Delta	% Attempt.	% Not Reached	SCB (N = 249) % Correct	Delta	% Attempt.	% Not Reached
1	81%	9.48	99%	0	85%	8.84	99%	0	61%	11.88	95%	0
2	78%	9.92	92%	0	84%	9.04	99%	0	65%	11.44	86%	0
3	59%	12.08	86%	0	82%	9.32	93%	0	39%	14.12	75%	0
4	69%	11.00	90%	0	90%	7.88	97%	0	56%	12.40	86%	0
5	76%	10.16	88%	0	87%	8.48	95%	0	66%	11.36	85%	0
6	76%	10.16	81%	0	90%	7.88	88%	0	59%	12.08	82%	0
7	60%	12.00	89%	0	80%	9.64	96%	0	24%	15.84	93%	0
8	73%	10.56	61%	0	92%	7.36	81%	0	60%	12.00	71%	0
9	32%	14.88	74%	0	68%	11.12	83%	0	14%	17.32	82%	0
10	73%	10.56	80%	0	85%	8.84	94%	0	60%	12.00	76%	0
11	51%	12.88	75%	0	70%	10.92	91%	0	45%	13.52	78%	0
12	59%	12.08	78%	2%	72%	10.68	90%	0	43%	13.72	80%	0
13	42%	13.80	54%	3%	68%	11.12	72%	0	22%	16.08	61%	1%
14	43%	13.72	86%	3%	65%	11.44	94%	0	35%	14.56	89%	1%
15	52%	12.80	46%	5%	75%	10.32	71%	0	32%	14.88	54%	3%
16	68%	11.12	74%	6%	85%	8.84	88%	1%	46%	13.40	77%	4%
17	40%	14.00	70%	9%	75%	10.32	83%	3%	24%	15.84	72%	6%
18	55%	12.48	58%	17%	70%	10.92	71%	6%	23%	15.96	64%	12%
19	47%	13.32	61%	23%	65%	11.44	80%	8%	36%	14.44	70%	14%
20	31%	15.00	38%	30%	56%	12.40	51%	13%	18%	16.68	43%	18%
21	21%	16.24	44%	32%	45%	13.52	61%	16%	7%	18.92	54%	19%
22	15%	17.16	56%	34%	33%	14.76	75%	19%	12%	17.68	55%	25%
23	14%	17.32	36%	46%	39%	14.12	65%	26%	7%	18.92	35%	31%
24	56%	12.40	19%	58%	59%	12.08	38%	41%	28%	15.32	27%	37%
25	5%	19.58	39%	61%	17%	16.80	53%	47%	6%	19.20	62%	38%

* % Correct is based on number attempting each item.
**Delta is the normal deviate, expressed in terms of a scale with mean of 13 and standard deviation of 4, which corresponds to the proportion of candidates attempting the item who answer it correctly.
***Not reached is the percent of people who did not attempt each item and did not attempt any subsequent items. For the last item it is simply the percent of people who did not respond to the item.

Form C

Item	RCB (N = 70) % Correct*	Delta**	% Attempt.	% Not*** Reached	RCW (N = 464) % Correct	Delta	% Attempt.	% Not Reached	SCB (N = 239) % Correct	Delta	% Attempt.	% Not Reached
1	78%	9.92	93%	0	83%	9.20	98%	0	62%	11.80	97%	0
2	74%	10.44	89%	0	90%	7.88	98%	0	63%	11.68	87%	0
3	62%	11.80	74%	0	79%	9.76	91%	0	28%	15.32	72%	0
4	68%	.2	84%	0	86%	8.68	97%	0	56%	12.40	86%	0
5	70%	92	86%	0	85%	8.84	95%	0	59%	12.08	85%	0
6	76%	.16	66%	0	90%	7.88	86%	0	57%	12.28	74%	0
7	57%	12.28	86%	0	82%	9.32	95%	0	26%	15.56	92%	0
8	77%	10.04	63%	0	90%	7.88	78%	0	63%	11.68	71%	0
9	33%	14.76	70%	1%	65%	11.44	82%	0	10%	18.12	83%	0
10	77%	10.04	76%	1%	84%	9.04	93%	0	53%	12.68	75%	0
11	52%	12.80	74%	3%	75%	10.32	83%	0	45%	13.52	80%	0
12	51%	12.88	73%	3%	73%	10.56	88%	0	29%	15.20	76%	2%
13	31%	15.00	56%	3%	67%	11.24	69%	1%	23%	15.96	57%	3%
14	39%	14.12	70%	4%	67%	11.24	92%	1%	30%	15.08	81%	5%
15	26%	15.56	39%	16%	73%	10.56	63%	2%	34%	14.68	51%	8%
16	67%	11.24	64%	19%	84%	9.04	82%	3%	45%	13.52	67%	10%
17	50%	13.00	54%	21%	71%	10.80	76%	5%	18%	16.68	55%	14%
18	38%	14.32	46%	27%	73%	10.56	62%	10%	27%	15.44	61%	18%
19	42%	13.80	47%	30%	68%	11.12	72%	12%	27%	15.44	55%	24%
20	33%	14.76	30%	36%	54%	12.60	43%	22%	19%	16.52	38%	28%
21	8%	18.64	36%	37%	47%	13.32	48%	24%	8%	18.64	47%	33%
22	8%	18.64	37%	46%	30%	15.08	63%	29%	10%	18.12	42%	39%
23	8%	18.64	36%	54%	34%	14.64	48%	42%	8%	18.64	32%	49%
24	33%	14.76	17%	69%	43%	13.72	21%	60%	28%	15.32	20%	58%
25	0	---	27%	73%	20%	16.36	35%	65%	5%	19.58	39%	61%

* % Correct is based on number attempting each item.

**Delta is the normal deviate, expressed in terms of a scale with mean of 13 and standard deviation of 4, which corresponds to the proportion of candidates attempting the item who answer it correctly.

***Not reached is the percent of people who did not attempt each item and did not attempt any subsequent items. For the last item it is simply the percent of people who did not respond to the item.